UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/626,875	07/25/2003	Francine R. Chen	A3053-US-NP XERZ 2 01564	1845

61962 7590 12/11/2009

FAY SHARPE / XEROX - PARC
1228 EUCLID AVENUE, 5TH FLOOR
THE HALLE BUILDING
CLEVELAND, OH 44115

EXAMINER	
BURKE, JEFF A	

ART UNIT	PAPER NUMBER
2159	

MAIL DATE	DELIVERY MODE
12/11/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 10/626,875
Filing Date: July 25, 2003
Appellant(s): CHEN ET AL.

Mark S. Svat
Reg. No. 34,261
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 9/10/2009 appealing from the Office action mailed 4/10/2009

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

No amendment after final has been filed.

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is substantially correct. The changes are as follows: Claims 77-80 are unpatentable under 35 U.S.C. 103(a) over Wayne Charles, "Topic Detection and Tracking in Chinese and

English" (hereinafter Wayne) in view of Ralph Brown, "Dynamic Stopwording for Story Link Detection" (hereinafter Brown).

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

6,606,620	**Sundaresan et al.**
5,835,905	**Pirolli et al.**
6,961,954	**Maybury et al.**

Wayne Charles, "Topic Detection and Tracking in English and Chinese", Published 2000.

Ralph Brown, Dynamic Stopwording for Story Link Detection, pp. 1-5.

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Rejections - 35 USC § 101

1. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

2. Claims 77-80 are rejected under 35 U.S.C. 101 as being directed to unpatentable subject matter.

3. With respect to claims 77-80, a method is recited. However, to be considered statutory subject matter, a method must either (1) be tied to another statutory class by specific recitation of the other statutory class, or (2) transform the underlying subject matter to a different state or thing. *Diamond v. Diehr*, 450 U.S. 175, 184 (1981); *Parker v. Flook*, 437 U.S. 584, 588 n.9 (1978); *Gottschalk v. Benson*, 409 U.S. 63, 70 (1972); *Cochrane v. Deener*, 94 U.S. 780, 787-88 (1876).

With respect to the first prong, these methods do not meet the requirements set forth because there is not a recitation of a particular apparatus within the meaning of the test. With respect to the second prong, the underlying subject matter is not transformed to a different state or thing, because it is merely electrical signals.

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. Claims 1-5,9-10, 14-,24, 28-29, 33-43, 47-48, 52-62, 66-67, 71-76 and 81-88 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sundaresan et al. (US Patent 6,606,620 B1) in view of Pirolli et al. (US 5,835,905) and further in view of Maybury et al. (US 6;961,954 B1).

As per claim 1 Sundaresan et al. is directed to a computer-implemented method of determining predictive models for a linked event detection system comprising the steps of: determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents" and see col. 6, lines 65-67 and col. 9, lines 9-17, *disclosing that web pages can have URLS which identify their source or a class label which can also identify the source of the document*); determining link label information for the at least one story-pair (column 9, lines 8-9); determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information (column 10, lines 5-13); and Sundaresan et al. does not teach determining inter-story similarity vectors for at least one story-pair. Pirolli et al. teaches determining inter-story similarity vectors for at least one story-pair (Pirolli et al., column 7, lines 53-65, wherein "pages" could mean "stories").

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Pirolli et al. to include determining inter-story similarity vectors for at least one story-pair because it provides the similarity measure of documents. Sundaresan et al. as modified still does not teach

the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event. Maybury et al. does teach the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al. as modified by teachings of Maybury et al. to include the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source identified stories are related to the same event because they indicate related information (Maybury et al., paragraph 16, line 33).

As per claim 2 Sundaresan et al. as modified is directed to a step of determining inter-story similarity vectors comprises the steps of: determining at least one inter-story similarity metric for the story-pairs (Sundaresan et al., column 4, lines 9-25); and determining at least one source-pair statistics for the at least one story-pair (Sundaresan et al., column 10, lines 15-17 and see Maybury, paragraph 16, lines 31-33, *disclosing that links are provided to determine relatedness of documents*).

As per claim 3 Sundaresan et al. as modified is directed to a determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics (Sundaresan et al., column 10, lines 17-22).

As per claim 4 Sundaresan et al. as modified is directed to a determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics (Sundaresan et al., column 10, lines 16-22).

As per claim 5 Sundaresan et al. as modified is directed to the inter-story similarity metric is normalized based on at least one of subtraction and division (Sundaresan et al., column 8, lines 22-27).

As per claim 9 Sundaresan et al. as modified is directed to a comprising the steps of transforming the source-identified training stories (Sundaresan et al., column 1, line 63, wherein the "training stories" are in English).

As per claim 10 Sundaresan et al. as modified is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming (Sundaresan et al., column 1, line 63; column 2, line 43, wherein the HTML and XML are in English, therefore translation will not be necessary).

As per claim 14 Sundaresan et al. as modified is directed to at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic (Sundaresan et al., column 10, lines 15-17).

As per claim 15 Sundaresan et al. as modified is directed to at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier (Sundaresan et al., column 3, lines 13-14).

As per claim 16 Sundaresan et al. as modified is directed to at least one of the source-pair similarity statistics are determined based on a source hierarchy (Sundaresan et al., column 3, lines 50-51).

As per claim 17 Sundaresan et al. as modified is directed to the source hierarchy is determined based on at least one source characteristic (Sundaresan et al., column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 18 Sundaresan et al. as modified is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic (Sundaresan et al., column 3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 19 Sundaresan et al. as modified is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristic of the new source (Sundaresan et al., column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 20 Sundaresan et al. is directed to a linked event detection training system comprising:

an input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have input/output device circuit);

a memory (column 7, lines 34-351 wherein it is inherent for computer to have memory);

a processor that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have a processor);

and a predictive model determining circuit that determines and stores at least one predictive model based on the inter-story similarity vectors and the link label information (column 10, lines 5-13)

Sundaresan et al. does not teach an inter-story similarity vector determining circuit that determines an inter-story similarity vectors in memory for at least one story-pair of the source-identified stories.

Pirolli et al. teaches an inter-story similarity vector determining circuit that determines an inter-story similarity vectors in memory for at least one story-pair of the source-identified stories (Pirolli et al., column 7, lines 53-65, wherein "pages" could mean "stories").

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Pirolli et al. to include an inter-story similarity vector determining circuit that determines an inter-story similarity vectors in memory for at least one story-pair of the source-identified stories because it provides the similarity measure of documents.

Sundaresan et al. as modified still does not teach the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event.

Maybury et al. does teach the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al. as modified by teachings of Maybury et al. to include the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the

linked source-identified stories are related to the same event because they indicate related information (Maybury et al., paragraph 16, line 33).

As per claim 21 Sundaresan et al. as modified is directed to the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the at least one story-pair (Sundaresan et al., column 4, lines 9-25);

and a similarity statistics determining circuit that determines at least one source-pair statistic for the at least one story-pair (Sundaresan et al., column 10, lines 15-17).

As per claim 22 Sundaresan et al. as modified is directed to the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics (Sundaresan et al., column 10, lines 17-22).

As per claim 23 Sundaresan et al. as modified is directed to the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics (Sundaresan et al., column 10, lines 16-22).

As per claim 24 Sundaresan et al. as modified is directed to at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation (.Sundaresan et al., column 8, lines 22-27).

As per claim 28 Sundaresan et al. as modified is directed to a comprising the step of transforming the source-identified training stories (Sundaresan et al., column 1, line 63, wherein the "training stories" are in English).

As per claim 29 Sundaresan et al. as modified is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming (Sundaresan et al., column 1, line 63; column 2, line 43, wherein the HTML and XML are in English therefore translation will not be necessary).

As per claim 33 Sundaresan et al. as modified is directed to the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic (Sundaresan et al., column 10, lines 15-17).

As per claim 34 Sundaresan et al. as modified is directed to the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier (Sundaresan et al., column 3, lines 13-14).

As per claim 35 Sundaresan et al. as modified is directed to the source-pair identified similarity statistic is determined based on a source hierarchy (Sundaresan et al., column 3, lines 50-51).

As per claim 36 Sundaresan et al. as modified is directed to the source hierarchy is determined based on at least one of a source characteristic (Sundaresan et al., column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 37 Sundaresan et al. as modified is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic (Sundaresan et al., column 3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 38 Sundaresan et al. as modified is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source (Sundaresan et al., column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 39 Sundaresan et al. is directed to a computer-implemented method of linked event detection comprising the steps of:

determining source-identified stories (column 3, lines 16-17, wherein "stories" means "documents");

determining at least one predictive model in the memory for link detection (column 10, lines 5-13);

and determining a link between the story-pairs based on the predictive model and the inter-story similarity vector (column 10, lines 5-13, wherein sorting determines the link); and

displaying the link on a computer or storing the link in an information repository (column 6, lines 57-59)

Sundaresan et al. does not teach determining inter-story similarity vectors in a memory for the story-pairs of the source-verified stories.

Pirolli et al. teaches determining inter-story similarity vectors in a memory for the story-pairs of the source-identified stories (Pirolli et al., column 7, lines 53-65, wherein "pages" could mean "stories").

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Pirolli et al. to include determining inter-story similarity vectors in a memory for the story-pairs of the source-verified stories because it provides the similarity measure of documents.

Sundaresan et al. as combined still does not teach the link indicating the story-pair are related to the same event.

Maybury et al. does teach the link indicating the story-pair are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al. as modified by teachings of Maybury et al. to include the link indicating the story-pair are related to the same event (Maybury et al., paragraph 16, line 33).

As per claim 40 Sundaresan et al. as modified is directed to a step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for each story-pair (Sundaresan et al., column 4, lines 9-25);

and determining source-pair statistics for the story-pairs (Sundaresan et al., column 10, lines 15-17).

As per claim 41 Sundaresan et al. as modified is directed to a determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics (Sundaresan et al., column 10, lines 17-22).

As per claim 42 Sundaresan et al. as modified is directed to a determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics (Sundaresan et al., column 10, lines 16-22).

As per claim 43 Sundaresan et al. as modified is directed to the inter-story similarity metric is normalized based on at least one of subtraction and division (Sundaresan et al., column 8, lines 22-27).

As per claim 47 Sundaresan et al. as modified is directed to a comprising the step of transforming the source-identified training stories (column 1, line 63, wherein the "training stories" are in English).

As per claim 48 Sundaresan et al. as modified is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming (Sundaresan et al., column 1, line 63; column 2, line 43, wherein the HTML and XML are in English therefore translation will not be necessary).

As per claim 52 Sundaresan et al. as modified is directed to the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic (Sundaresan et al., column 10, lines 15-17).

As per claim 53 Sundaresan et al. as modified is directed to the at least one predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes-classifier (Sundaresan et al., column 8, lines 22-27).

As per claim 54 Sundaresan et al. as modified is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source (Sundaresan et al., column 3, lines 50-51).

As per claim 55 Sundaresan et al. as modified is directed to the source hierarchy is determined based on at least one of a source characteristic (Sundaresan et al., column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 56 Sundaresan et al. as modified is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic (Sundaresan et al., column 3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 57 Sundaresan et al. as modified is directed to the source-pair similarity statistic for a new source is determined based on at least one source

characteristics of the new source (Sundaresan et al., column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 58 Sundaresan et al. is directed to linked event detection system comprising:

an input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have input/output device circuit);

a memory (column 7, lines 34-35, wherein it is inherent for computer to have memory);

a processor that receives source-identified training stories via the input/output circuit (column 7, lines 34-35, wherein it is inherent for computer to have processor);

and a link determining circuit that determines and displays on a computer or stores in an information repository, links between story-pairs based on a predictive model in the memory and the inter-story similarity vectors (column 10, lines 5-13, wherein sorting determines the link; column 6, lines 57-59).

Sundaresan et al. does not teach an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories.

Pirolli et al. teaches an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for the story-pairs of the source-

identified stories (Pirolli et al., column 7, lines 53-65, wherein "pages" could mean "stories").

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Pirolli et al. to include an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories because it provides the similarity measure of documents.

Sundaresan et al. as combined still does not teach the link indicating the story-pair are related to the same event.

Maybury et al. does teach the link indicating the story-pair are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al. as modified by teachings of Maybury et al. to include the link indicating the story-pair are related to the same event (Maybury et al., paragraph 16, line 33).

As per claim 59 Sundaresan et al. as modified is directed to the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the story-pairs (Sundaresan et al., column 4, lines 9-25);

and a similarity statistics determining circuit that determines source-pair statistics for the story-pairs (column 10, lines 15-17).

As per claim 60 Sundaresan et al. as modified is directed to the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics (.Sundaresan et al., column 10, lines 17-22).

As per claim 61 Sundaresan et al. as modified is directed to the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics (Sundaresan et al., column 10, lines 16-22).

As per claim 62 Sundaresan et al. as modified is directed to at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation (Sundaresan et al. column 8, lines 22-27).

As per claim 66 Sundaresan et al. as modified is directed to a comprising the step of transforming the source-identified training stories (Sundaresan et al., column 1, line 63, wherein the "training stories" are in English).

As per claim 67 Sundaresan et al. as modified is directed to transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming (Sundaresan et al., column 1, line 63; column 2, line 43, wherein the HTML and XML are in English therefore translation will not be necessary).

As per claim 71 Sundaresan et al. as modified is directed to the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic (Sundaresan et al., column 10, lines 15-17).

As per claim 72 Sundaresan et al. as modified is directed to the predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes classifier (Sundaresan et al., column 8, lines 22-27).

As per claim 73 Sundaresan et al. as modified is directed to the source-pair identified similarity. statistic is determined based on a source hierarchy (Sundaresan et al., column 3, lines 50-51).

As per claim 74 Sundaresan et al. as modified is directed to the source hierarchy is determined based on at least one of a source characteristic (.Sundaresan et al., column 3, lines 61-65, wherein "characteristic" means "leaf").

As per claim 75 Sundaresan et al. as modified is directed to the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic (Sundaresan et al., column 3, lines 54-60, wherein "language characteristic" means how the words in a document are related to each other).

As per claim 76 Sundaresan et al. as modified is directed to the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source (Sundaresan et al.., column 3, lines 50-53, wherein each new source has different hierarchy).

As per claim 81 Sundaresan et al. is directed to computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to determine at least one predictive model for a linked event detection system by executing steps comprising:

determining source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

determining link label information for the at least one story-pair (column 9, lines 8-9);

and determining and storing at least one predictive model in the memory based on the inter-story similarity vector and the link label information (column 7, lines 24-25; column 10, lines 5-13);

Sundaresan et al. does not teach determining inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories

Pirolli et al. teaches determining inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories (Pirolli et al., column 7,lines 53-65, wherein "pages" could mean "stories").

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Pirolli et al. to include determining inter-story similarity vectors in a memory for at least one story-pair of the

source-identified training stories because it provides the similarity measure of documents.

Sundaresan et al. as combined still does not teach the link label information indicating training stories are related to the same event.

Maybury et al. does teach the link label information indicating training stories are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al. as modified by teachings of Maybury et al. to include the link label information indicating training stories are related to the same event (Maybury et al., paragraph 16, line 33).

As per claim 82 Sundaresan et al. is directed to computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to determine at least one predictive model for a linked event detection system, the computer readable program code comprising:

instructions to determine source-identified training stories (column 3, lines 16-17, wherein "stories" means "documents");

instructions to determine link label information for the at least one story-pair (column 9, lines 8-9);

and instructions to determine and store at least one predictive model in the memory based on the inter-story similarity vector and the link label information (column 7, lines 24-25; column 10, lines 5-13); and

Sundaresan et al. does not teach instructions to determine inter-story similarity vectors in memory for at least one story-pair of the source-identified training stories.

Pirolli et al. teaches instructions to determine inter-story similarity vectors in memory for at least one story-pair of the source-identified training stories (Pirolli et al., column 7, lines 53-65, wherein "pages" could mean "stories").

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Pirolli et al. to include instructions to determine inter-story similarity vectors in memory for at least one story-pair of the source-identified training stories because it provides the similarity measure of documents.

Sundaresan et al. as combined still does not teach the link label information indicating training stories are related to the same event.

Maybury et al. does teach the link label information indicating training stories are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al. as modified by teachings of

Maybury et al. to include the link label information indicating training stories are related to the same event (Maybury et al., paragraph 16, line 33).

As per claim 83 Sundaresan et al. is directed to computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to detect linked events by executing steps comprising:

determining source-identified stories (column 3, lines 16-17, wherein "stories" means "documents");

determining at least one predictive model in the memory for link detection (column 9, lines 8-9);

determining a link between story-pairs based on the at least one predictive model and the inter-story similarity vectors (column 10, lines 5-13); and

displaying the link on a computer or storing the link in .an information repository, (column 6, lines 57-59).

Sundaresan et al. does not teach determining inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories.

Pirolli et al. teaches determining inter-story, similarity vectors in a memory for the at least one story-pair of the source-identified stories (Pirolli et al., column 7, lines 53-65, wherein "pages" could mean "stories"). It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al.

by teachings of Pirolli et al. to include determining inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories because it provides the similarity measure of documents.

Sundaresan et al. as combined still does not teach the link indicating the story-pairs are related to the same event.

Maybury et al. does teach the link indicating the story-pairs are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al.. as modified by teachings of Maybury et al. to include the link indicating the story-pairs are related to the same event (Maybury et al., paragraph 16, line 33).

As per claim 84 Sundaresan et al. is directed to computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code executable to program a computer to detect linked events comprising the steps of:

instructions to determine source-identified stories(column 3, lines 16-17, wherein "stories" means "documents");

instructions to determine at least one predictive model in a memory for link detection (column 9, lines 8-9);

instructions to determine a link between story-pairs based on the predictive model and the inter-story similarity vectors (column 10, lines 5-13);); and

instructions displaying the link on a computer or storing the link in an information repository, (column 6, lines 57-59).

Sundaresan et al. does not teach instructions to determine inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories.

Pirolli et al. teaches instructions to determine inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories (Pirolli et al., column 7, lines 53-65, wherein "pages" could mean "stories").

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to modify Sundaresan et al. by teachings of Pirolli et al. to include instructions to determine inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories because it provides the similarity measure of documents.

Sundaresan et al. as combined ,still does not teach the link indicating the story-pairs are related to the same event.

Maybury et al. does teach the link indicating the story-pairs are related to the same event (paragraph 16, lines 31-33).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the Sundaresan et al. as modified by teachings of

Maybury et al. to include the link indicating the story-pairs are related to the same event (Maybury et al., paragraph 16, line 33).

As per claims 85 and 86 Sundaresan et al. as modified is directed to determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric (Sundaresan et al., column 3, lines 25-29, wherein the statistical algorithm uses metric for the computations).

As per claims 87 and 88 Sundaresan et al. as modified is directed to at least one of the predictive models is a trained predictive model (Sundaresan et al., column 10, lines 29-33, wherein the "trained predictive model" is determined by use of statistical model).

6. Claims 6-8, 25-27, 44-46, and 63-65 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sundaresan et al. (US Patent 6,606,620 B1) in view of Pirolli et al. (US 5,835,905) and further in view of Gange et al. (US 2004/006559 A1) and further in view of Maybury et al. (US 6,961,954 B1).

As per claims 6, 25, 44 and 63 Sundaresan et al. as modified fails to teach the use of probability based metric and a Euclidean based similarity metric.
Gan.qe et al. teaches the use of Euclidean distance (Gan.qe et al., page 3, paragraph 0045).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to further modify Sundaresan et al. as modified by teachings of Ganqe et al. to include the use of Euclidean distance as it is metrics often used in the database field to compute distances between similar terms.

As per claims 7, 26, 45 and 64 Sundaresan et al. as modified fails to teach the use of similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric. Ganqe et al: teaches the use of Tanimoto coefficient (Ganqe et al., page 3, paragraph 0045).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to further modify Sundaresan et al. as modified by teachings of Ganqe et al. to include the use of Tanimoto coefficient as it is metrics often used in the database field to compute distances between similar terms.

A per claims 8, 27, 46 and 65 Sundaresan et al. as modified fails to teach the use of inter-story similarity metric is a cosine-distance based metric. Ganqe et al. teaches the use of Cosine coefficient (Ganqe et al., page 3, paragraph 0045).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to further modify Sundaresan et al. as modified by teachings of Ganqe et al. to include the use of Cosine coefficient as it is metrics often used in the database field to compute distances between similar terms.

7. Claims 11-13, 30-32, 49-51 and 68-70 are rejected under 35 U.S.C. 103(a) as being unpatentable over Sundaresan et al. (US Patent 6,606,620 B1) in view of Pirolli et al. (US 5,835,905) and in further view Zhou (US 2004/0002849 A1) and further in view of Maybury et al. (US 6,961,954 B1).

As per claims 11, 30, 49 and 68 Sundaresan et al. as modified fails to teach the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models. Zhou teaches the use of frequency-inverse (Zhou, page 3, column 2, paragraph 0030, lines 9-11).

It would have been obvious to one in of ordinary skill in the art at the time the invention was made to further modify Sundaresan et al. as modified by teachings of Zhou to include the use of frequency-inverse because it predicts effective example of sentence retrieval as stated on page 1, column 1, paragraph 0005 of Zhou.
As per claims 12, 37, 50 and 69 Sundaresan et al. as modified fails to teach the terms in source-identified term frequency-inverse story frequency models are based on language.

Zhou teaches that the retrieved samples are to aid in writing or translation (Zhou, page 3, paragraph 0030, lines 2-4, wherein writing or translating has basis in language). It would have been obvious to one in of ordinary skill in the art at the time the invention was made to further modify Sundaresan et al. as modified by teachings of Zhou to include the inverse-frequency based on language because term comparison includes terms of a language.

As per claims 13, 32, 51 and 70 Sundaresan et al. as modified fails to teach determining terms comprises the steps: determining a reference language; and determining reference language and non-reference language terms.

Zhou teaches the changing of sample terms from one mode to another (Zhou, page 3, paragraph 0032).

It would have been obvious to on of ordinary skill in the art at the time the invention was made to further modify Sundaresan et al. as modified by teachings of Zhou to include the determination of reference language since the correct translation requires the correct reference language.

8. Claims 77-80 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wayne (Topic Detection and Tracking in English and Chinese, Charles Wanye, Published in 2000) in view of view of Brown (Dynamic Stopwording for Story Link Detection, Ralph Brown).

As per claim 77, Wayne teaches a method of determining a stop word list comprising the steps of:

determining a source-identified training corpus of text information (pg. 166, section 3, and table 1, sources);

determining a verified first source-mode transformation of the source-identified training corpus text from a first mode to a second mode based on at least one of a verified transcription and a verified translation (pg. 166-167, sections 3.2-3.5, verified translations and transcriptions of the source corpora);

determining an un-verified second source-mode transformation of the source-identified training corpus text from a first mode to a second mode (pg. 166-167, sections 3.2-3.5, verified translations and transcriptions of the source corpora, specifically audio transcription to manual closed caption quality, but not verified as stated in the specification);

determining at least one transformation error associated with distribution differences between the first and second transformations and identified sources (Figure 5, determining the transformation errors, and distribution differences between Chinese audio/text and English audio/text);

Wayne does not explicitly disclose determining and storing at least one source-specific transformation action for the determined transformation errors in a memory and identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

Brown discloses determining and storing at least one source-specific transformation action for the determined transformation errors in a memory (page 2, column 1, lines 2-6); and identifying and transforming transformation errors in other

transformed source-identified texts based on the source-specific transformation actions in the memory (page 2, column 1, lines 2-7).

Wayne and Brown are analogous art because they are relevant to solving the same problem, transformation and translation errors in documents. It would have been obvious to one of ordinary skill in the art, at the time of the invention to modify the method of Wayne to include the functionality of Brown to allow documents to be auto-corrected. The suggestion/motivation to combine is that this process enhances the efficiency and effectiveness of the correction system. Therefore it would have been obvious to combine the above references to obtain the instant invention.

As per claim 78 Wayne as modified is directed to the first mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source (pg. 166, section 3.2).

As per claim 79 Wayne as modified is directed to the second mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source (pg. 166, section 3.2).

As per claim 80 the combination of Wayne and Wu do not disclose wherein the source-specific transformation is at least one of a removal, a repair and a normalization transformation. Brown is directed to wherein the source-specific transformation is at

least one of a removal, a repair and a normalization transformation (Brown, page 2, column 1, lines 4-6).

Wayne, Wu and Brown are analogous art because they are relevant to solving the same problem, transformation and translation errors in documents. It would have been obvious to one of ordinary skill in the art, at the time of the invention to modify the method of Wayne and Wu to include the transformation of Brown to allow one of the above transformation methods to be used. The suggestion/motivation to combine is to allow efficient removal of stopwords (Brown, page 2, column 1, lines 4-6). Therefore it would have been obvious to combine the above references to obtain the instant invention.

(10) Response to Argument

With respect to claim 1, Applicant argues that the cited art of record fails to teach or disclose "source identified training stories", the examiner respect fully disagree. Sundaresan discloses the creation of classifications of web pages, i.e. stories (col. 7, lines 54-61) and it is clear that these documents are source identified. Source identified, as interpreted by the examiner means that the stories contain a source, i.e. a location. All files contain a file name, which is a location, specifically web pages contain a URL, or domain name, which is the source of the web page. Sundaresan makes this clear at col. 6, lines 65-67, ***explicitly stating that web sites correspond to a particular***

internet domain name*, and again at col. 6, line37-39, *a URL is a unique address that fully specifies the location of a content object on the internet, i.e. its source.

Furthermore, applicant contends that the art fails to utilization of the source information. The examiner disagrees for two reasons:

(1) Sundaresan clearly discloses creating a structured vector for each document based on extracting textual and structural information from the document (col. 10, lines 5-13), in this case structural information can read on the URL of the document, and

(2) The claimed limitations of claim 1 **do not seem to disclose the utilization of the source-identification, therefore the applicant's is arguing limitations not found in the claim language itself. The claim merely recites the determination of source-identified training stories and does nothing further with the source information.**

Applicant further contends that the cited art of record fails to teach or disclose "inter-story similarity vectors for source-identified stories", the examiner respectfully disagrees. First, as to the source-identification information, **no where in the claim language is it claimed that the source information is used in the determination.** Secondly, Pirolli clearly discloses the creation of inter-document similarity vectors at col. 7, lines 53-67, ***for each web page a document similarity is determined using the dot product of the similarity vectors computed for the documents, this is done for each pair of pages.***

Applicant further contends that the cited art of record fails to teach "link label information", the examiner respectfully disagrees. Sundaresan discloses this limitation at least at col. 9, lines 8-9, ***comparing label information of the document to class labels.*** Furthermore, Sundaresan goes on to detail that some documents have class labels and that these labels are fed to the classifier to learn term distribution for the labels.

Applicant finally contends that the cited art of record fails to teach "that the documents are linked to the same event", the examiner respectfully disagrees. Maybury clearly discloses this limitation at col. 16, lines 31-33, ***disclosing that links are provided between various records, i.e. documents or web pages, and the story segment to which they pertain, i.e. events, to permit the retrieval of related data.***

In summary, the combination of Sudaresan, Pirolli and Maybury, taken for what they fairly disclose, teach all of the claimed, and unclaimed but argued limitations of claim 1. The primary art teaches a system for determining classification of documents, which are source identified, the secondary art modifies that classification by providing a means to determine similarity between document, and the tertiary document modifies the art by providing that events are a valid way of determining relatedness of documents.

With respect to claim 2, Applicant argues that the cited art of record fails to teach or disclose "that the statistical information is not based on source identification data", the examiner disagrees for at least two reasons:

(1) The claimed limitations of claim 1 **do not seem to disclose the utilization of the source-identification as statistics,** but merely determines a statistics of the source-pair.

(2) Furthermore, Sundaresan still meets this limitation. As seen at col. 9, lines 18-27, ***documents may already have a class label associated with the, i.e. the label can be thought of as a source, and further at*** col. 10, lines 15-22***, statistical information that will be useful in classifying documents is collected, and used as statistics and Maybury has already disclosed that the event information is useful in determining relatedness of events.*** Therefore, the combination discloses the claimed limitation.

Claims 21, 40, 59, 3-19, 22-38, 41-57, 60-76 and 85-88 have been argued based on their dependence from claim 1.

With respect to claim 77, applicant argues the rejection of the claim under 35 U.S.C. 101 as being patent eligible subject matter, the examiner respectfully disagrees. The claim recites a method, however, in order for a method to be patentable, under current U.S.P.T.O. guidelines the method must be (1) tied to a particular machine or apparatus or (2) transform the underlying subject matter to a different state or thing.

In this method there is no recitation of a machine or apparatus, much less a particular machine or apparatus, therefore the initial inquiry fails. Furthermore, there is no transformation of the underlying subject matter to a different state or thing. The

underlying subject matter is merely data, more specifically it is an electrical signal which is used in a computation to create more data, or electrical signals.

Finally, with respect to claim 77, applicant argues that the cited art fails to teach or disclose "transforming errors in source-identified texts", the examiner respectfully disagrees. The applicant relies on the interpretation that the transformations must occur in the body of the text. However, the examiner suggests that another, reasonable interpretation exists, and that is that the transformations merely occur, it is irrelevant where they occur according to the claim. Nowhere in the claim language does the claim require the transformation to occur in the "actual" text or the "body" of the text. Brown teaches the limitation as interpreted by the examiner at pg. 2, left column, ***stopwords are effectively removed from each the documents, i.e. the stopwords are the transformation errors, which are then transformed in other texts.***

In summary, the examiner's interpretation of the limitation is reasonable, and Brown meets the limitation by removing stopwords from the texts as required by the claim language.

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

/JEFFREY BURKE/

Examiner, Art Unit 2159

Conferees:

/Luke S. Wassum/
Primary Examiner, Art Unit 2167

/James Trujillo/
Supervisory Patent Examiner, Art Unit 2159